

数量化4類の基準による展開法における

制約条件について¹⁾

岡 本 安 晴

- | | |
|----------------------|------------------|
| 0. はじめに | 3. シミュレーションによる比較 |
| 1. 数量化4類の基準による展開法の解法 | 4. 雑 考 |
| 2. 布置の分散の扱い方について | |

0. はじめに

空間内の幾つかの点に対して、それらの座標値が与えられれば点間の距離を求めることができる。逆に、それらの点間の距離が与えられたときに、座標値を求めようとするのが多次元尺度法 (MDS) である。データとして距離そのものではなく距離と何らかの関係にあるもの、特に距離と単調増加/減少の関係にあるものが与えられているときには、与えられたデータと単調関係にある距離の推定が行われる。このタイプの MDS はノンメトリックと呼ばれている。このノンメトリック MDS に対して、距離の推定を行わず、単に与えられたデータに示される大まかな関係を表す点の布置を求めようとする方法が数量化4類である。数量化の考案者林によれば、要素(点) i と j の類似度(距離と単調減少の関係にある)データ e_{ij} が与えられたとき、 e_{ij} 一型数量化(数量化4類)は次のように理解されるべきであるという。

「漠としたものを漠として扱い、その情報のもとで、大局的に要素の空間布置やグルーピングを行ってみることにその特色があり、あまり e_{ij} の数値に鋭く反応しないようにするところに e_{ij} 一型数量化の特色があるというべきである。」(林・鮑戸(1976), p.22)。

数量化4類では、 e_{ij} が与えられたとき、次の目的関数 Q を最大にするように点 i の座標値 $x(i)$ が求められる。

$$Q = - \sum e_{ij} \cdot \{x(i) - x(j)\}^2 \quad (0.1)$$

多次元尺度法あるいは数量化4類では、点間の距離(あるいは、それと単調関係にある非類似度や類似度)が、基本的にはすべての点間にわたって与えられている。これに対し、点が

1) 計算において、京都大学大型計算機センター、及び金沢大学総合情報処理センターを利用した。プログラムは FACOM OS IV/F4 MSP PASCAL で書かれているが、興味のある人は筆者までお問い合わせ下さい(電子メール宛先 C00279@jpnac.bitnet 又は C00279@sinet.ad.jp)。

2つのグループ $\{x_i\}_{i=1,\dots,M}$ と $\{y_j\}_{j=1,\dots,N}$ に分けられていて、これら2つのグループ間でのみ距離 $\{d_{ij} = x_i \text{ と } y_j \text{ の距離}\}_{i=1,\dots,M; j=1,\dots,N}$ が与えられているときに $\{x_i\}$ と $\{y_j\}$ の布置を求める方法は展開法と呼ばれている。岡本(1991)は(0.1)式の基準を展開法の場合に適用した分析例を報告した。展開法の場合には(0.1)式は次の(0.2)式の形に書ける。

$$Q = - \sum_{i=1}^M \sum_{j=1}^N e_{ij} \cdot \{x(i) - y(j)\}^2 \quad (0.2)$$

ここに、 $x(i)$ 、 $y(j)$ は点 x_i 、 y_j の座標値である。(0.2)式は、(0.1)式において欠損値の含まれる特殊な場合である。

1. 数量化4類の基準による展開法の解法

(0.2)式に対する解を数量化4類の方法(cf.岩坪(1987))に倣って求める。

まず

$$\mathbf{x}' = [x(1), \dots, x(M)], \mathbf{y}' = [y(1), \dots, y(N)]$$

$$E = (e_{ij}): \quad e_{ij} \text{ を } (i, j) \text{ 要素とする } M \times N \text{ 行列}$$

$$E_1 = \text{diag}(e_{i\cdot}), \quad e_{i\cdot} = \sum_j e_{ij}$$

$$E_2 = \text{diag}(e_{\cdot j}), \quad e_{\cdot j} = \sum_i e_{ij}$$

とおけば、(0.2)式は次のように書ける。

$$Q_1 = -\mathbf{x}'E_1\mathbf{x} + 2\mathbf{x}'E\mathbf{y} - \mathbf{y}'E_2\mathbf{y}$$

さらに、以下の議論では

$$e_{ij} \geq 0, \quad e_{i\cdot} > 0, \quad e_{\cdot j} > 0$$

であるとする。

\mathbf{x} と \mathbf{y} は、 Q_1 を最大にする値として求められるが、 $Q_1 \leq 0$ であるので、 \mathbf{x} と \mathbf{y} が 0 に近づく程 Q_1 は大きくなる。相対的な布置を与える座標値として \mathbf{x} と \mathbf{y} を求めるので、 \mathbf{x} と \mathbf{y} の分散に次の制約を設ける。

$$\mathbf{x}'E_1\mathbf{x} + \mathbf{y}'E_2\mathbf{y} = 2 \quad (1.1)$$

このとき、 Q_1 は次のようになる。

$$Q_1 = -2 + 2\mathbf{x}'E\mathbf{y}$$

(1.1)式の制約条件下での Q_1 の最大化(極大化)を考えるので、ラグランジュ未定係数 λ を用いた次の目的関数 R を考える。

$$R = 2\mathbf{x}'E\mathbf{y} - \lambda(\mathbf{x}'E_1\mathbf{x} + \mathbf{y}'E_2\mathbf{y} - 2)$$

R の偏導関数を求めると

$$\frac{\partial R}{\partial \mathbf{x}} = 2E\mathbf{y} - 2\lambda E_1\mathbf{x}$$

$$\frac{\partial R}{\partial \mathbf{y}} = 2E'\mathbf{x} - 2\lambda E_2\mathbf{y}$$

$$\frac{\partial R}{\partial \lambda} = -\mathbf{x}'E_1\mathbf{x} - \mathbf{y}'E_2\mathbf{y} + 2$$

これらの偏導関数を 0 とおいて、

$$E\mathbf{y} = \lambda E_1\mathbf{x} \quad (1.2)$$

$$E'\mathbf{x} = \lambda E_2\mathbf{y} \quad (1.3)$$

(1.2) と (1.3) より

$$\lambda \mathbf{x}'E_1\mathbf{x} = \mathbf{x}'E\mathbf{y} = \lambda \mathbf{y}'E_2\mathbf{y}$$

故に、(1.1) より

$$\lambda \mathbf{x}'E_1\mathbf{x} = \lambda \mathbf{y}'E_2\mathbf{y} = \lambda$$

(1.2) より

$$\mathbf{x}'E\mathbf{y} = \lambda \mathbf{x}'E_1\mathbf{x} = \lambda \quad (1.4)$$

故に、これらの $\mathbf{x}, \mathbf{y}, \lambda$ に対して

$$Q_1 = -2 + 2\lambda$$

$Q_1 \leq 0$ より、

$$\lambda \leq 1$$

Q_1 ができるだけ大きくなるように λ を選ぶが、 $\lambda = 1$ に対しては

$$\mathbf{x} = (1, \dots, 1)' = \mathbf{1}_M, \quad \mathbf{y} = (1, \dots, 1) = \mathbf{1}_N$$

1 が M 個

1 が N 個

が解となる。すなわち

$$E\mathbf{1}_N = \mathbf{1} \cdot E_1\mathbf{1}_M$$

$$E'\mathbf{1}_M = \mathbf{1} \cdot E_2\mathbf{1}_N$$

これらは、 $\{e_{ij}\}$ に対する $\{x_{il}\}$ と $\{y_{jl}\}$ の布置としては不適であるので、 $\lambda < 1$ に対する解を求める。

$$\mathbf{u} = E_1^{-1/2} \mathbf{x}, \quad \mathbf{v} = E_2^{-1/2} \mathbf{y}, \quad G = E_1^{-1/2} \cdot E \cdot E_2^{-1/2}$$

とおけば、(1.2) 式と (1.3) 式は次のようになる。

$$G\mathbf{v} = E_1^{-1/2} \cdot E \cdot E_2^{-1/2} \cdot E_2\mathbf{y} = E_1^{-1/2} \cdot E\mathbf{y} = \lambda E_1^{-1/2} \mathbf{x} = \lambda \mathbf{u} \quad (1.2)'$$

$$G'\mathbf{u} = E_2^{-1/2} \cdot E' \cdot E_1^{-1/2} \cdot E_1\mathbf{x} = E_2^{-1/2} \cdot E'\mathbf{x} = \lambda E_2^{-1/2} \mathbf{y} = \lambda \mathbf{v} \quad (1.3)'$$

(1.2)' と (1.3)' は G の特異値分解 (SVD) を与える式である。すなわち、 G の SVD を求める

ることにより、 $\lambda, \mathbf{x}, \mathbf{y}$ が求まる。

$\lambda \neq 0$ であれば、(1.4)より、

$$\mathbf{x}'E_1\mathbf{x}=\mathbf{y}'E_2\mathbf{y}=1 \quad (1.5)$$

を得る。これは、 \mathbf{x} と \mathbf{y} の全体の分散に対する制約(1.1)式が \mathbf{x} と \mathbf{y} の個々の互いに等しい分散に対する制約に分解されることを示している。しかし、この制約は Q_1 を最大(より正確には極大)にする0でない \mathbf{x} と \mathbf{y} を求めるために便宜上置かれたものである。

(0.2)式は1次元解の形でたてられているが、2次元以上の布置が求められているときには、 λ を $0 < \lambda < 1$ の範囲で大きいものから順に選び、それらに対応する \mathbf{x} と \mathbf{y} がとられる。 p 次元解に対するものを $\{\lambda_s, \mathbf{x}_s, \mathbf{y}_s\}; s=1, \dots, p$ とおくとき、上の分散に対する制約条件がどのように解に反映させられるべきであるかという問題が提起される。制約 $\mathbf{x}_s'E_1\mathbf{x}_s=\mathbf{y}_s'E_2\mathbf{y}_s=1$ は、解法上便宜的に設定されたものであるので、多次元解においては、これら各次元における分散の次元間をどのように扱うかという問題が生じる。ここでは次の3つの考え方を扱うが、これらの問題は高根芳雄、丘本正の両氏の私信による岡本の発表(1991)に対するコメントを参考にして扱われており、ここに謝意を表する次第である。

2. 布置の分散の扱い方について

分散の扱い方として次の3つの考え方を取り上げる。

考え方. 1

制約条件(1.5)は、0でない解を求めるために便宜上設定されたものであるが、解の制約条件には違いがないので、そのまま用いる。これ以上は、分散の次元間について考えない。従って、 p 次元空間における p 個の次元に対する解は、全て(1.5)式を満たすようにとられる。

考え方. 2

制約条件(1.5)を、0でない解を求めるための便宜的なものと考え、分散の次元間関係は次のように考える。

G の特異値分解を次のようにおく。

$$G=U\Lambda V'$$

$$U=(\mathbf{u}_1, \dots, \mathbf{u}_r), \quad V=(\mathbf{v}_1, \dots, \mathbf{v}_r)$$

$$\Lambda=\text{diag}(\lambda_1, \dots, \lambda_r)$$

ただし、 r は G のランクである。

いま、

$$X=(\mathbf{x}_1, \dots, \mathbf{x}_r), \quad Y=(\mathbf{y}_1, \dots, \mathbf{y}_r)$$

とおけば

$$U=E_1^{1/2}X, \quad V=E_2^{1/2}Y$$

ゆえに

$$G = E_1^{1/2} X \Lambda Y' E_2^{1/2}$$

$$E_1^{-1} E E_2^{-1} = X \Lambda Y'$$

行列 H を

$$H = (h_{ij}), \quad h_{ij} = e_{i \cdot} e_{j \cdot} / (e_{i \cdot} e_{j \cdot})$$

とおけば

$$H = X \Lambda Y'$$

すなわち

$$h_{ij} = \sum_s \lambda_s x_{is} y_{js}$$

但し、

$$\mathbf{x}'_s = (x_{1s}, \dots, x_{ms}), \quad \mathbf{y}'_s = (y_{1s}, \dots, y_{ns})$$

いま

$$A = (a_{ik}) = X \Lambda X'$$

$$B = (b_{jl}) = Y \Lambda Y'$$

とおけば

$$a_{ik} = \sum_s \lambda_s x_{is} x_{ks}$$

$$b_{jl} = \sum_s \lambda_s y_{js} y_{ls}$$

故に

$$d_{ij}^2 = a_{ii} - 2h_{ij} + b_{jj}$$

とおけば

$$d_{ij}^2 = \sum_s \lambda_s (x_{is}^2 - 2x_{js} y_{is} + y_{js}^2)$$

$$= \sum_s (\sqrt{\lambda_s} \cdot x_{is} - \sqrt{\lambda_s} \cdot y_{js})^2 \quad (2.1)$$

(2.1)式は、数量化4類における考え方(岩坪(1987)、(6.26)式))に倣ったものである。点 i と j の第 s 次元における布置をそれぞれ $\sqrt{\lambda_s} \cdot \mathbf{x}_s$ と $\sqrt{\lambda_s} \cdot \mathbf{y}_s$ で与えると、 r 次元空間における i と j の距離は d_{ij} となる。

d_{ij}^2 と e_{ij} の関係は次のようである。

$$e_{ij} = \left[\left\{ d_{ij}^2 \times \left(-\frac{1}{2} \right) + \frac{a_{ii} + b_{jj}}{2} \right\} \right] \times e_{i \cdot} \times e_{j \cdot}$$

よって、点 i あるいは j が原点から遠ざかっているほど、類似性は強調されることになる。
 $e_{i \cdot}$ 及び $e_{j \cdot}$ は行及び列のスケーリング因子として作用している。

上の解釈により、 $X \Lambda^{1/2}$ 及び $Y \Lambda^{1/2}$ を E に対応する布置と考えることの1つの妥当性

が与えられる。さらに、

$$(\sqrt{\lambda_s} \mathbf{x}_s)' E_1 (\sqrt{\lambda_s} \mathbf{x}_s) = (\sqrt{\lambda_s} \mathbf{y}_s)' E_2 (\sqrt{\lambda_s} \mathbf{y}_s) = \lambda_s$$

すなわち、 $\sqrt{\lambda_s} \mathbf{x}_s$ 及び $\sqrt{\lambda_s} \mathbf{y}_s$ の分散が λ_s であることにより、 λ_s の大きさから布置の実質的な次元数を決めることの根拠も与えられる。

考え方. 3

Q_1 を最大(極大)にする \mathbf{x} と \mathbf{y} を Takane(1980) 及び Takane *et al.* (1991) における方法に倣って求める。

まず、与えられた \mathbf{x} に対して、 Q_1 を最大にする \mathbf{y} を求める。

$$\frac{\partial Q_1}{\partial \mathbf{y}} = -2E'\mathbf{x} + 2E_2\mathbf{y}$$

$\partial Q_1 / \partial \mathbf{y} = 0$ とおけば

$$E'\mathbf{x} = E_2\mathbf{y} \quad (2.2)$$

(2.2) 式により与えられる \mathbf{y} を Q_1 に代入することにより Q_1 は \mathbf{x} の関数 $Q_2(\mathbf{x})$ となる。

すなわち

$$\begin{aligned} Q_2(\mathbf{x}) &= -\mathbf{x}'E_1\mathbf{x} + 2\mathbf{x}'E \cdot E_2^{-1} E'\mathbf{x} - (E_2^{-1} E'\mathbf{x})' E_2 (E_2^{-1} E'\mathbf{x}) \\ &= -\mathbf{x}'E_1\mathbf{x} + 2\mathbf{x}'E \cdot E_2^{-1} E'\mathbf{x} - \mathbf{x}'E \cdot E_2^{-1} E'\mathbf{x} \\ &= -\mathbf{x}'E_1\mathbf{x} + \mathbf{x}'E \cdot E_2^{-1} E'\mathbf{x} \end{aligned}$$

$Q_2(\mathbf{x})$ も \mathbf{x} が 0 に近づくとき、その最大値 0 に近づくので、 $Q_2(\mathbf{x})$ の最大(極小)化を \mathbf{x} の分散についての制約条件 $\mathbf{x}'E_1\mathbf{x} = 1$ のもとで考える。ラグランジュ未定係数 μ を用いて次の目的関数 $R_2(\mathbf{x})$ をおく。

$$R_2(\mathbf{x}) = \mathbf{x}'E E_2^{-1} E'\mathbf{x} - \mu(\mathbf{x}'E_1\mathbf{x} - 1)$$

偏導関数を求めると

$$\frac{\partial}{\partial \mathbf{x}} R_2(\mathbf{x}) = 2E E_2^{-1} E'\mathbf{x} - 2\mu E_1\mathbf{x}$$

$\frac{\partial}{\partial \mathbf{x}} R_2(\mathbf{x}) = \mathbf{0}$ とおいて

$$E E_2^{-1} E'\mathbf{x} = \mu E_1\mathbf{x}$$

(2.2) より

$$E\mathbf{y} = \mu E_1\mathbf{x} \quad (2.3)$$

(2.3) 式の μ は (1.4)、(1.5) 式における λ と $\mu = \lambda^2$ の関係にある。又、この「考え方. 3」の解法における \mathbf{x} と \mathbf{y} の分散は

$$\mathbf{x}'E_1\mathbf{x} = 1, \quad \mathbf{y}'E_2\mathbf{y} = \mu = \lambda^2 \quad (2.4)$$

多次元解において、「考え方. 1」の場合と同様に (2.4) 式の分散をそのまま布置の分散として採用すれば、 \mathbf{x} の分散は各次元にわたって全く等しく 1 であるが、 \mathbf{y} の分散は次元に対応する G の特異値 λ の平方に等しくなる。

尚、 $\mathbf{x}'E_1\mathbf{x}=1$ の制約条件で求められた「考え方. 3」の解 $\mathbf{x}, \mathbf{y}, \mu=\lambda^2$ に対して

$$\begin{aligned} Q_1 &= -1 + 2\mu - \mu \\ &= -1 + \lambda^2 \quad (=P_2 \text{ とおく}) \end{aligned}$$

他方、第1章の $\mathbf{x}'E_1\mathbf{x}=1, \mathbf{y}'E_2\mathbf{y}=1$ の制約条件で求められた解に対しては

$$Q_1 = -2 + 2\lambda \quad (=P_1 \text{ とおく})$$

従って

$$\begin{aligned} P_2 - P_1 &= -1 + \lambda^2 + 2 - 2\lambda \\ &= (1 - \lambda)^2 > 0 \end{aligned}$$

すなわち、 $\mathbf{x}'E_1\mathbf{x}=1$ の制約条件のみの方がより大きな値 $P_2(>P_1)$ が得られているが、このとき

$$\mathbf{x}'E_1\mathbf{x} + \mathbf{y}'E_2\mathbf{y} = 1 + \lambda^2$$

上の条件、すなわち

$$\mathbf{x}'E_1\mathbf{x} = \mathbf{y}'E_2\mathbf{y} = \frac{1}{2}(1 + \lambda^2)$$

の制約条件で解を求め Q_1 の値 P_3 を計算すれば

$$\begin{aligned} P_3 &= -(1 + \lambda^2) + 2\lambda \frac{1}{2}(1 + \lambda^2) \\ &= -1 - \lambda^2 + \lambda + \lambda^3 \end{aligned}$$

P_2 と比べると

$$\begin{aligned} P_3 - P_2 &= -1 - \lambda^2 + \lambda + \lambda^3 - (-1 + \lambda^2) \\ &= \lambda(1 - \lambda)^2 > 0 \end{aligned}$$

すなわち、 P_2 より大きい値 P_3 が得られる。

以上、 \mathbf{x} と \mathbf{y} の布置における分散の扱い方について3つの考え方を取り上げたが、これらの *simulation* による比較を行ったものを次に報告する。

3. シミュレーションによる比較

数量化の方法は、布置の正確な再現を目的とするものではなく、データの示す大まかな傾向を集約して表すところにその目的があることを思い起こせば、細かい比較は無益である。又、次章で取り上げるように、数量化4類の基準による展開法の解にはその基準(0.2)式の意味と解法による特有の制約(歪み)がある。パターンの再現性を厳しくいうのならば、それに応じたモデルが用いられるべきである。しかし、この再現性を目指したモデルにおいて、座標値の推定のアプローチに初期値が必要とされている場合には、local minimum に陥らないようにできるだけ良い初期値を用いるのが望ましい。又、予めデータの示す大まかな

傾向を数量化の方法で把握しておくことは、後の再現性を目指したモデルによる分析の際にも有用である。岡本(1992)は、SD法データの分析において、まず「考え方、2」の方法による解を求め、最尤法による解法の初期値とするとともに、最尤法により得られた解のGram-Schmidtの直交化法による回転を行うときの基底の選定における情報としても活用している。従って、前章の3つの考え方による解の布置パターンの再現性の比較を行っておくことは有益である。

まず、点が格子状に配置されている規則的な場合の比較を行う。これにより、解の歪み具合が明瞭になる。次に、より「生」のデータに近い場合として、布置がランダムに与えられた場合の比較が行われる。以後、点の2つのグループ $\{x_i\}$ と $\{y_j\}$ は被験者(Subject)および対象(Object)のグループと呼ばれる。これは、展開法が、心理学において、個人(被験者)の対象に対する好みの評定が、その個人の理想の対象と評定対象との距離によって決まるとするモデル(Coombs(1964))に始まっていることによる。

1. 格子状布置の場合

48個の対象を 8×6 の格子状に配し、その間に35の理想点(被験者)を 7×5 の格子状に置く。

$$\text{対象 } j \text{ の座標} = (y_{j1}, y_{j2}) = (((j-1) \div 8) + 0.5, ((j-1) \bmod 8) + 0.5)$$

$$\text{被験者 } i \text{ の座標} = (x_{i1}, x_{i2}) = (((i-1) \div 7) + 1, ((i-1) \bmod 7) + 1)$$

である。これらを図1及び図2に図示する。これらの図では、中央で交差している「一」及び「I」の連なりで表される直線がそれぞれ第1次元および第2次元の座標軸である。従って、中央の交点は原点(0,0)である。対象と被験者の布置が表示されているこれらの図における座標軸の単位は共通であるので、図を重ね合わせることでにより対象と被験者の同時分布が得られる。展開法の解としてはこの対象と被験者を同じ図に表示する方が適切であるが、図が煩雑になって解釈の妨げとなる嫌いがある。この章では図示は全て対象と被験者が別の図として表示されるが、原点と座標軸の単位を共通にすることにより、同時分布も読み取れるようにする。座標軸の単位は、 $\max_{i,j,s} \{|x_{is}|, |y_{js}|\}$ の大きさの位置が外側の正方形の枠の一文字分内側になるようにとられている。この単位の大さは全ての次元にわたって共通である。図中、対象 j 及び被験者 i は、それぞれ $j=1$ から26まで、及び $i=1$ から26までは英大文字のAからZで表され、 $j>26$ 及び $i>26$ については疑問符?で表されている。

格子状に設定された被験者 i と対象 j の座標に対して、類似度 e_{ij} を次式で与える。

$$e_{ij} = \max_e - d_{ij} \quad (3.1)$$

ここに

$$d_{ij} = \sqrt{(x_{i1} - y_{j1})^2 + (x_{i2} - y_{j2})^2}$$

$$\max_e = \max_{i,j} \{d_{ij}\}$$

INITIAL OBJECT CONFIGURATION

X - 軸 → 次元 1
Y - 軸 → 次元 2

	H	P	X	?	?	?
	G	O	W	?	?	?
	F	N	V	?	?	?
	E	M	U	?	?	?
	D	L	T	?	?	?
	C	K	S	?	?	?
	B	J	R	Z	?	?
	A	I	Q	Y	?	?

図 1 格子状布置の場合の対象の分布。 $y_{js} > 0$ であるので第 I 象限のみに配置されている。中央の交点が原点であり、座標軸の単位は座標の絶対値のすべての次元及び対象と被験者(図 2 のもの)にわたる最大値が外枠から一文字内側の位置になるようにとられている。図 2 と座標軸の単位が同じであるので、両図を重ねることにより対象と被験者の同時分布が得られる。

INITIAL SUBJECT CONFIGURATION

X - 軸 → 次元 1
Y - 軸 → 次元 2

	G	N	U	?	?
	F	M	T	?	?
	E	L	S	Z	?
	D	K	R	Y	?
	C	J	Q	X	?
	B	I	P	W	?
	A	H	O	V	?

図 2 格子状布置の場合の被験者の分布。 $x_{is} > 0$ であるので第 I 象限のみに配置されている。

である。

この e_{ij} に対して、第2章の3つの考え方による解の布置の比較を行う。「考え方1」による布置を図3及び図4、「考え方2」によるものを図5及び図6に示す。「考え方3」によるものは、 $\mathbf{x}'E_1\mathbf{x}=1$ (図では $U'U=1$ で表されている) の制約のみによる布置を図7と図8に、 $\mathbf{y}'E_2\mathbf{y}=1$ (図では $V'V=1$) の制約のみによるものを図9と図10に分けて示されている。中央の交点が原点であること、座標軸の単位が対象と被験者の座標値の絶対値の最大値を基に設定されていることは図1と図2の場合と同じである。座標軸の単位の設定は、図3と図4、図5と図6、図7と図8、図9と図10の各ペアごとに独立に行われている。図1から図10までを比較すると、「考え方3」による布置において、対象と被験者の布置を重ねてみるとそれぞれの分布の広がり著しい差異の認められることが分かる。分布の中心が原点にあることについては次章で取り上げるが、展開法においては布置の相対的な位置関係のみが意味をもつことに注意されたい。「考え方1」と「考え方2」では目立った差異はないが、「考え方2」による布置の方が元の布置(図1及び図2)の矩形状に忠実である。

布置の再現性を量的に評価するために、始めに設定された格子状の布置(図1及び図2)を *target* として、それぞれの「考え方」による解に拡大縮小と重心の移動を許す直交プロクラステス回転 (Schönemann and Carroll (1970)) を適用した。プロクラステス回転において、被験者と対象は同一の空間内の布置 (joint map) としてまとめて扱うことにする。回転後の

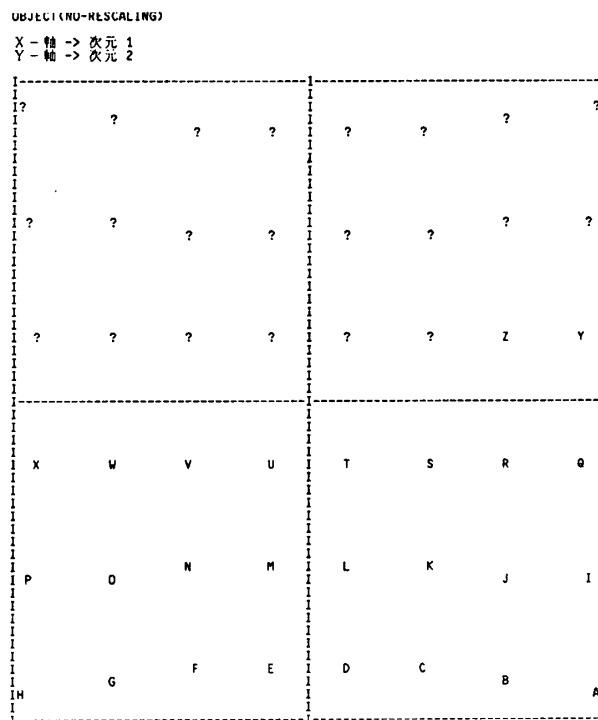


図3 「考え方1」による対象の布置。図4と重ねることにより被験者との同時分布が得られる。

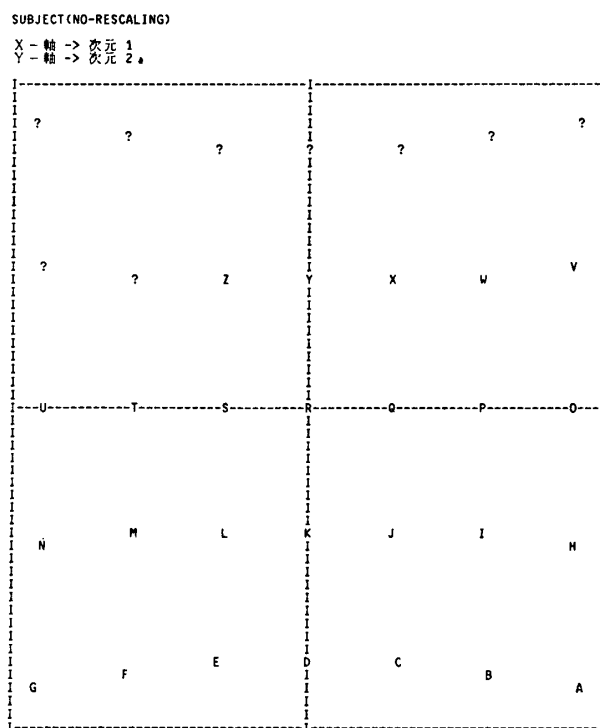


図4 「考え方. 1」による被験者の布置。

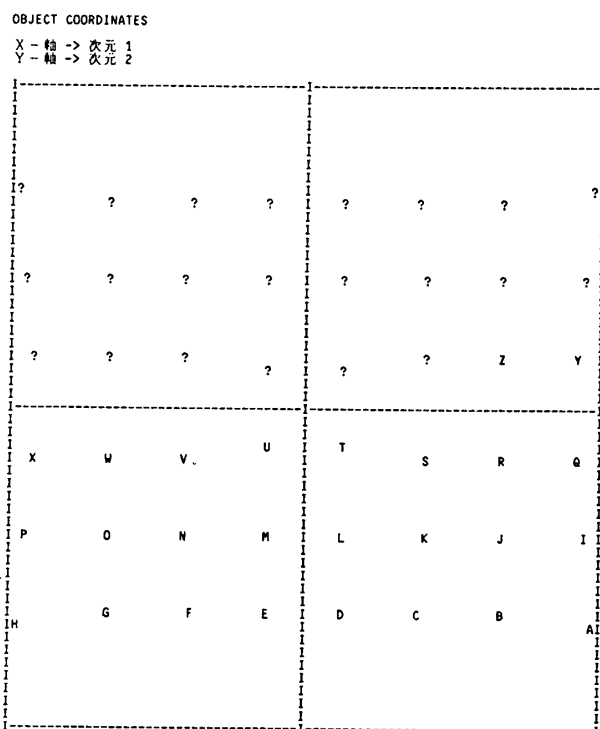


図5 「考え方. 2」による対象の布置。図6と重ねることにより被験者との同時分布が得られる。

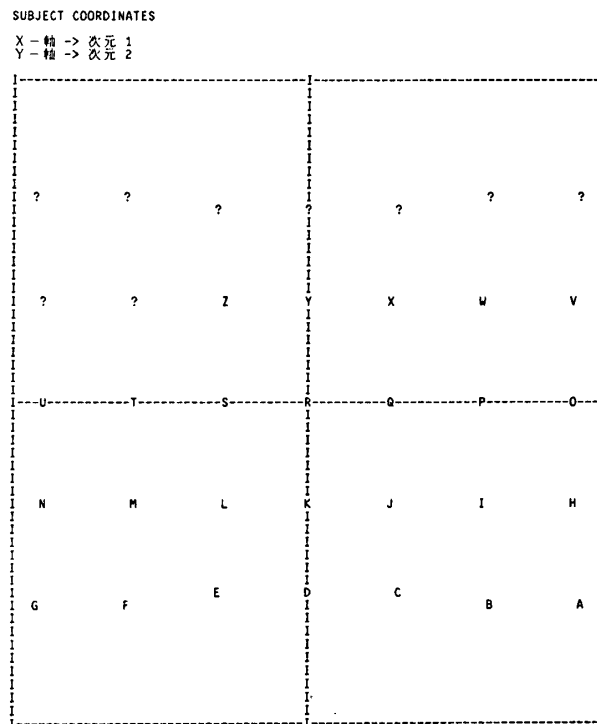


図6 「考え方. 2」による被験者の布置。

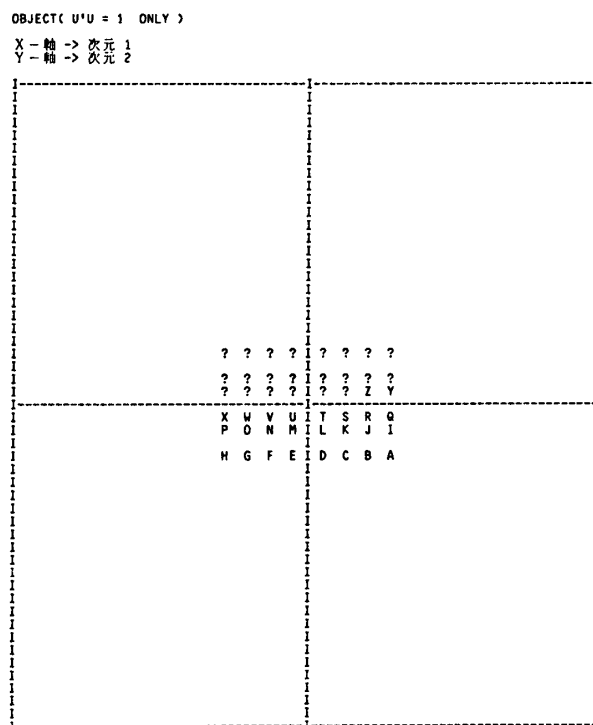


図7 「考え方. 3 ($U^*U = 1$ の制約のみ)」による対象の布置。図8と重ねることにより被験者との同時分布が得られる。

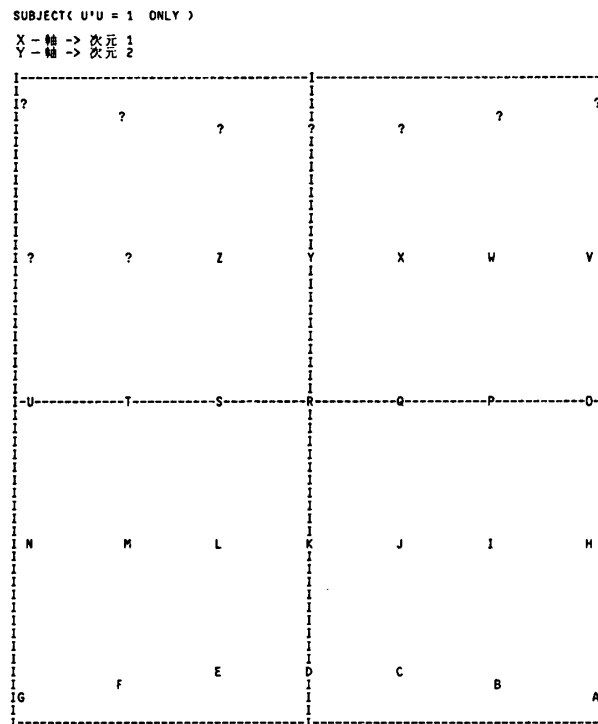


図8 「考え方. 3 ($U'U = 1$ の制約のみ)」による対象の布置。

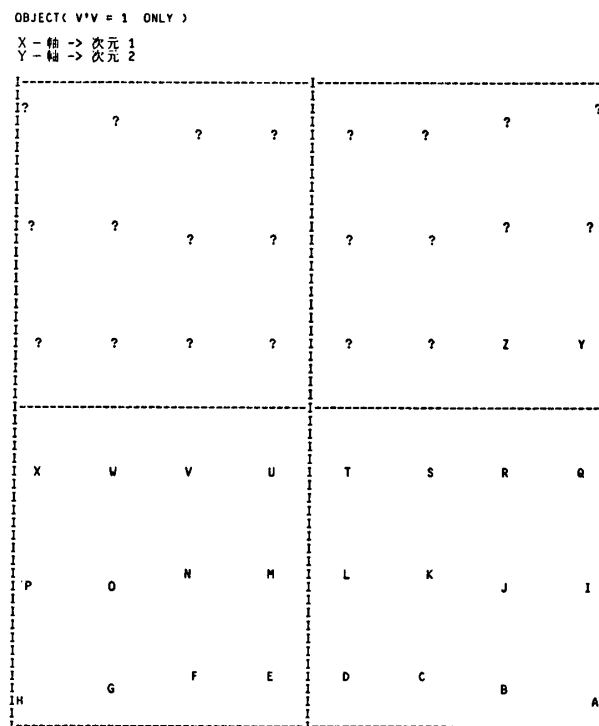


図9 「考え方. 3 ($V'V = 1$ の制約のみ)」による対象の布置。図10と重なることにより被験者との同時分布が得られる。

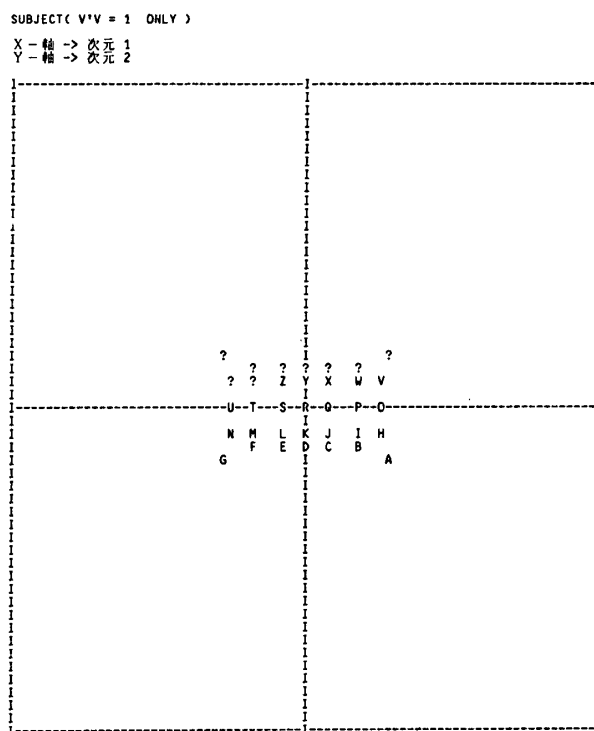


図10 「考え方. 3 ($V'V = 1$ の制約のみ)」による被験者の布置。

布置と target の布置とを比べ、その誤差の2乗和を求め、表1にそれらの値を示した。「考え方. 2」による解の再現性が一番優れ、「考え方. 3」によるものは非常に悪いことが分かる。

解 法	誤差の2乗和
考え方. 1	15.69
考え方. 2	3.25
考え方. 3 ($u'u = 1$)	237.33
考え方. 3 ($v'v = 1$)	115.71

表1 格子状布置に対する3つの「考え方」による再現性。それぞれの解を原点移動とスケーリングを伴う直交プロクラステス回転によりできるだけ最初の格子状布置に近づけた後、その差異を座標値の差の2乗和のすべての対応する点にわたる総和として求めた。

2. ランダム布置の場合

次に、より実データに近い場合として点の布置がランダムに与えられたものに対して比較を行う。M個の理想点(被験者)とN個の対象の2次元及び3次元空間における座標値を一樣乱数により与え、これらの座標値を有する被験者*i*と対象*j*の類似度 e_{ij} を格子状布置の場合と同様に(3.1)式により設定する。3次元の布置においては、 d_{ij} は3次元ユークリ

ッド空間における距離である。これらの e_{ij} に対して、それぞれの「考え方」による布置を求め、その再現性の評価のために、それらの布置をプロクラステス回転したものと元の布置との差の2乗和を求める。さらに、「考え方. 2」に対する差の2乗和 D_2 を基準にして、「考え方. 1」、「考え方. 3 ($U'U=1$)」、「考え方. 3 ($V'V=1$)」に対する差の2乗和 D_1 、 D_{3U} 、 D_{3V} の比率 D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 を計算する。MとNの各組合わせに対して、上の e_{ij} データの作成と各「考え方」の比較 D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 の計算を100回繰り返した。こうして得られた D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 のそれぞれ100個の値の集合について、それらの25パーセンタイル (P_{25})、50パーセンタイル (中央値M)、75パーセンタイル (P_{75}) を求め、MとNの各組合わせ別に2次元空間の場合を表2に、3次元空間の場合を表3に示した。表

M	N		D_1/D_2	D_{3U}/D_2	D_{3V}/D_2
10	10	P_{25}	1.11	4.04	4.58
		M	1.23	5.93	6.00
		P_{75}	1.44	7.79	8.62
100	10	P_{25}	0.76	2.00	12.09
		M	0.96	2.94	18.33
		P_{75}	1.15	4.30	25.85
20	10	P_{25}	1.05	4.35	7.82
		M	1.24	6.57	12.36
		P_{75}	1.74	9.76	17.16
100	50	P_{25}	1.06	21.14	39.09
		M	1.17	32.12	56.56
		P_{75}	1.46	42.42	81.62
100	100	P_{25}	1.09	44.00	45.64
		M	1.21	63.37	61.73
		P_{75}	1.45	104.90	98.93

表2 2次元空間に一樣乱数によりランダムに設定されたM個の理想点(被験者)とN個の対象の布置に対する類似度データ e_{ij} の場合における各「考え方」による解の比較。MとNの各組合わせについて、布置の乱数による設定と e_{ij} の算出、それらの与えられた e_{ij} に対する各「考え方」による解の布置の計算とプロクラステス回転後の誤差の2乗和 (D_1 = 「考え方. 1」のもの、 D_2 = 「考え方. 2」のもの、 D_{3U} = 「考え方. 3 ($U'U=1$ の制約)」のもの、 D_{3V} = 「考え方. 3 ($V'V=1$ の制約)」のもの)、及び D_2 を基準とした比率 D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 の算出ということをして100回繰り返し、得られた D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 それぞれ100個の値の集合について25パーセンタイル (P_{25})、50パーセンタイル (中央値M)、75パーセンタイル (P_{75}) を求めた。

2 および表3において比 D_1/D_2 、 D_{3U}/D_2 、 D_{3V}/D_2 の値が、($M=100$, $N=10$)の組合わせに対する D_1/D_2 の場合を除いて1より大きいことから、「考え方. 2」の方法による布置の再現性が最も優れていることが分かる。($M=100$, $N=10$)の組合わせに対する D_1/D_2 の

M	N		D_1/D_2	D_{3U}/D_2	D_{3V}/D_2
10	10	P_{25}	1.28	3.93	3.83
		M	1.47	4.82	4.81
		P_{75}	1.67	5.86	6.30
100	10	P_{25}	0.76	1.64	10.50
		M	0.85	2.13	13.59
		P_{75}	0.97	2.90	18.42
20	10	P_{25}	1.13	3.40	6.77
		M	1.34	4.60	8.59
		P_{75}	1.51	6.23	11.26
100	50	P_{25}	1.12	17.04	34.27
		M	1.27	22.78	43.18
		P_{75}	1.47	30.63	57.79
100	100	P_{25}	1.19	33.50	36.58
		M	1.31	48.54	48.40
		P_{75}	1.72	65.47	65.45

表3 3次元空間に一様乱数によりランダムにM個の理想点(被験者)とN個の対象の布置を設定して各「考え方」を比較したもの。3次元空間であることを除けば表2における場合と同じ方法である。

M値0.96と0.85も $1 \div 0.96 \doteq 1.04$ 及び $1 \div 0.85 \doteq 1.18$ であるから、表2及び表3全体の値と比べるとこの値は「考え方、2」と「考え方、1」の再現性の(M=100, N=10)における大きな優劣を示すものではない。MとNの値が大きくなると「考え方、3」の布置の再現性は非常に悪くなっているが、特に2次元の場合(表、2)にその傾向が著しい。

4. 雑 考

数量化4類を批判的に検討しているものとして丘本ら(1973)及びSaito (1982)のものがあがあるが、数量化4類の方法による展開法も目的関数(0.2)式とその解法による特性がある。

Gの特異値分解((1.2)'式、(1.3)'式)により解を求めるとき、 $\lambda_0=1$ に対する解 $u_0=E_1^{1/2} \mathbf{1}_M$, $v_0=E_2^{1/2} \mathbf{1}_N$ が除かれている。このため、第s次元に対する解 $x_s=E_1^{-1/2} u_s$ と $y_s=E_2^{-1/2} v_s$ は u_0 と v_0 に対して次の関係にある。

$$u_s' u_0 = 0, \quad v_s' v_0 = 0$$

すなわち

$$x_s' E_1 \mathbf{1}_M = 0 \quad y_s' E_2 \mathbf{1}_N = 0$$

$$\sum_i e_i x_{is} = 0 \quad \sum_j e_j y_{js} = 0$$

第 s 次元における布置 \mathbf{x}_s と \mathbf{y}_s は共にそれぞれの布置の中心が原点にある。数量化 4 類に倣った展開法におけるこの $\{\mathbf{x}_i\}$ の布置の中心が $\{\mathbf{y}_j\}$ の布置の中心に一致するという強い制約は Schönemann(1970)の展開法における解法では見られない。彼の解法では $\{\mathbf{x}_i\}$ と $\{\mathbf{y}_j\}$ の間の距離データ $\{d_{ij}\}$ が与えられたとき、 $\{\mathbf{x}_i\}$ と $\{\mathbf{y}_j\}$ の布置の中心が異なった位置にあるものとして行列 $\left(-\frac{1}{2} \cdot d_{ij}^2\right)$ の中心化が行 ($\{\mathbf{x}_i\}$) と列 ($\{\mathbf{y}_j\}$) について別々に行われているので、解の \mathbf{x}_s と \mathbf{y}_s それぞれの布置の中心が一致するという制約はない。しかし、数量化 4 類に倣った方法の方が Schönemann の方法より簡便であり、又、 d_{ij} が誤差を含まない正確な距離データでなければならないという制約もない。

目的関数(0.2)式は $\mathbf{e} = (e_{ij})$ と $\mathbf{d}^{(2)} = (d_{ij}^2)$ の内積になっている。ここで、 \mathbf{e} と $\mathbf{d}^{(2)}$ はそれぞれ e_{ij} と d_{ij} を適当な順序で並べた $M \times N$ 次元のベクトルである。

$Q = -\mathbf{e} \cdot \mathbf{d}^{(2)}$ の最大化は $\mathbf{d}^{(2)}$ の長さが一定という条件のもとでは $-\mathbf{e}$ と $\mathbf{d}^{(2)}$ の角度 θ の最小化を意味する。 $-\mathbf{e} = c \cdot \mathbf{d}^{(2)}$, $c =$ 正定数のときに $\theta = 0$ となるが、これは e_{ij} が正の値をとる類似度データであれば不可能である。 \mathbf{x}_i と \mathbf{y}_j の真の距離を δ_{ij} とするとき

$$e_{ij} = -a\delta_{ij}^2 + b \quad a, b > 0$$

とおけば

$$\begin{aligned} Q &= -\sum e_{ij} d_{ij}^2 \\ &= a \sum \delta_{ij}^2 d_{ij}^2 - b \sum d_{ij}^2 \end{aligned}$$

$\sum d_{ij}^2 \doteq$ 定数とみなせば、 Q の最大化を行うことは 2 つのベクトル (δ_{ij}^2) と (d_{ij}^2) の間の角度の最小化を行うことになり、この意味で Q の最大化の妥当性が理解される。

(0.2)式によく似た基準 $\rho(X)$ が de Leeuw (1977) で扱われている。彼は $\rho(X)$ とともに次の基準 $\sigma(X)$ を挙げ、それらの関係を論じている。

$$\sigma(X) = \sum w_{ij} (\delta_{ij} - d_{ij}(X))^2 \quad (4.1)$$

$$\rho(X) = \sum w_{ij} \delta_{ij} d_{ij}(X) \quad (4.2)$$

$\sigma(X)$ は距離データ δ_{ij} と布置行列 X による距離 $d_{ij}(X)$ との差の重み w_{ij} 付 2 乗和であり、 $\rho(X)$ は(0.2)式と同じくベクトル (δ_{ij}) とベクトル $(d_{ij}(X))$ との重み付内積である。彼は、 $\sigma(X)$ の最小化と $\rho(X)$ の最大化とは本質的に同じであることを示しているが、この場合 δ_{ij} は類似度データではなく距離(非類似度)データであることに注意しなければならない。

e_{ij} として度数データ f_{ij} を考えた場合(0.2)式は Takane (1980) 及び Heiser (1987) の展開法における目的関数となる。

Takane は、M個の対象をN人の被験者にクラスター分けさせた場合のデータを展開法の考え方で分析する方法を述べている。対象 i の座標を x_i 、被験者 k におけるクラスター r の座標を $y_{r,k}$ とおき、 x_i と $y_{r,k}$ の距離を $d_{ir(k)}$ と表す。和 g を、対象 i が被験者 k によりクラスター r に分類されている場合のみ $d_{ir(k)}^2$ を和 g に加えたものとする。

すなわち

$$g = \sum_{\substack{k,i,j \\ k \text{ が } i \text{ を } r \text{ に分類}}} d_{ir(k)}^2 \quad (4.3)$$

i が r に分類されたとき $e_{ir(k)} = 1$ 、それ以外は $e_{ir(k)} = 0$ とすれば

$$Q = - \sum e_{ir(k)} d_{ir(k)}^2 = -g$$

Q の最大化は g の最小化となる。

Heiser (1987) は度数データ f_{ij} の展開法による分析を correspondence analysis の一方法として論じている。de Leeuw (1977) の $\sigma(X)$ ((4.1)式)において、度数 f_{ij} が正のときは距離データ $\delta_{ij} = 0$ 重み $w_{ij} = f_{ij}$ とし、度数 $f_{ij} = 0$ のときは $\delta_{ij} = 1$ 、 $w_{ij} = 0$ とおけば

$$\sigma(X) = \sum f_{ij} d_{ij}(X)^2$$

となり、これは Takane の目的関数 g ((4.3)式)と同じである。この Heiser の場合 $\rho(X) = 0$ である。

correspondence analysis において、データ行列 E に対する chi square distance なるものが論じられている (cf. Carroll et al. (1986)、Greenacre (1984))。

$$\overrightarrow{e_{i\cdot}} = \frac{1}{e_{i\cdot}} (e_{i1}, \dots, e_{iN})$$

とおく。すなわち

$$\begin{bmatrix} \overrightarrow{e_{1\cdot}} \\ \vdots \\ \overrightarrow{e_{M\cdot}} \end{bmatrix} = E_1^{-1} E$$

この E_1 により標準化された E の第 i 行 $\overrightarrow{e_{i\cdot}}$ と第 h 行 $\overrightarrow{e_{h\cdot}}$ の E_2^{-1} を重みとする距離を計算する。

$$E_1^{-1} E E_2^{-1} = X \Lambda Y'$$

であるから

$$\begin{bmatrix} \overrightarrow{e_{1\cdot}} \\ \vdots \\ \overrightarrow{e_{M\cdot}} \end{bmatrix} E_2^{-1} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} \Lambda Y'$$

ここに

$$\mathbf{x}_i = [x_{i1}, \dots, x_{ir}]$$

よって

$$\overrightarrow{e_{i\cdot}} - \overrightarrow{e_{h\cdot}} = (\mathbf{x}_i - \mathbf{x}_h) \Lambda Y' E_2$$

これより $\overrightarrow{e_{i\cdot}}$ と $\overrightarrow{e_{h\cdot}}$ の重み付き距離は

$$\begin{aligned} (\overrightarrow{e_{i\cdot}} - \overrightarrow{e_{h\cdot}}) E_2^{-1} (\overrightarrow{e_{i\cdot}} - \overrightarrow{e_{h\cdot}})' &= (\mathbf{x}_i - \mathbf{x}_h) \Lambda Y' E_2 E_2^{-1} E_2 Y \Lambda (\mathbf{x}_i - \mathbf{x}_h)' \\ &= (\mathbf{x}_i - \mathbf{x}_h) \Lambda Y' E_2 Y \Lambda (\mathbf{x}_i - \mathbf{x}_h)' \\ &= (\mathbf{x}_i \Lambda - \mathbf{x}_h \Lambda) (\mathbf{x}_i \Lambda - \mathbf{x}_h \Lambda)' \end{aligned}$$

上式は、第 i 行と第 h 行の重み付き距離は第 s 次元の解を特異値 λ_s でスケーリングした布置 $X\Lambda$ により与えられることを示している。

引用文献

- Carroll, J. D., Green, P. E. and Schaffer, C. M. (1986) Interpoint distance comparisons in correspondence analysis. *Journal of Marketing Research*, X X III, 271-280.
- Coombs, C. H. (1964) A theory of data. John Wiley and Sons, Inc.
- de Leeuw, J. (1977) Application of convex analysis to multidimensional scaling. In: J. R. Barra et al. (eds.). Recent development in statistics. North Holland Publishing Company, 133-145.
- Greenacre, M. J. (1984) Theory and applications of correspondence analysis. Academic Press, Inc.
- 林知己夫・鮑戸 弘(共編纂)(1976) 多次元尺度解析法—その有効性と問題点—, サイエンス社
- Heiser, W. J. (1987) Joint ordination of species and sites: The unfolding technique. In: P. Legendre and L. Legendre(eds.). Developments in numerical ecology. Springer-Verlag.
- 岩坪秀一(1987) 数量化法の基礎 朝倉書店
- 丘本 正・戸田 準(1973) 数量化理論 4 類と主座標分析法, 日本統計学会誌3, 41-53
- 岡本安晴(1991) 数量化Ⅳ類の基準による展開法, 日本行動計量学会第19回大会発表論文抄録集.
- 岡本安晴(1992) SD法への展開法の適用について 日本心理学会第56回大会発表論文抄録集.

- Saito, T. (1982) Contributions to e_{ij} -type quantification and a new method of multidimensional scaling. *Behaviormetrika*, No.12, 63-83.
- Schönemann, P. H. (1970) On metric multidimensional unfolding. *Psychometrika*, 35, 349-366.
- Schönemann, P. H. and Carroll, R. M. (1970) Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35, 245-255.
- Takane, Y. (1980) Analysis of categorizing behavior by a quantification method. *Behaviormetrika*, No.8, 75-86.
- Takane, Y., Yanai, H. and Mayekawa, S. (1991) Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, 56, 667-684.